

PCT

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE  
Bureau international

DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

<b>(51) Classification internationale des brevets <sup>5</sup> :</b> C12N 15/15, 15/62, C07H 21/04 C07K 13/00, C12P 21/00 // A61K 37/02 (C12N 15/15 C12R 1/865)	<b>A1</b>	<b>(11) Numéro de publication internationale:</b> WO 90/13646
<b>(21) Numéro de la demande internationale:</b> PCT/FR90/00306		<b>(43) Date de publication internationale:</b> 15 novembre 1990 (15.11.90)
<b>(22) Date de dépôt international:</b> 27 avril 1990 (27.04.90)	<b>(74) Mandataire:</b> WARCOIN, Jacques; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).	
<b>(30) Données relatives à la priorité:</b> 89/05687 28 avril 1989 (28.04.89) FR	<b>(81) Etats désignés:</b> AT (brevet européen), BE (brevet européen), CA, CH (brevet européen), DE (brevet européen), DK (brevet européen), ES (brevet européen), FR (brevet européen), GB (brevet européen), IT (brevet européen), JP, LU (brevet européen), NL (brevet européen), SE (brevet européen), US.	
<b>(71) Déposant (pour tous les Etats désignés sauf US):</b> TRANSGENE S.A. [FR/FR]; 16, rue Henri-Régnauld, F-92400 Courbevoie (FR).	<b>Publiée</b> <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des revendications, sera republiée si de telles modifications sont requises.</i>	
<b>(72) Inventeurs; et</b> <b>(75) Inventeurs/Déposants (US seulement) :</b> ACHSTETTER, Tilman [DE/DE]; Uhlandweg 11, D-7602 Oberkirch (DE). NGUYEN, Martine [FR/FR]; 21, rue du Paradis, F-67670 Wittersheim (FR). LEMOINE, Yves [FR/FR]; 4, rue des Alisiers, F-67100 Strasbourg (FR). REICH-HART, Jean-Marc [FR/FR]; 34, rue de Rotterdam, F-67000 Strasbourg (FR).		
<b>(54) Title:</b> APPLICATION OF NOVEL DNA FRAGMENTS AS A CODING SEQUENCE FOR A SIGNAL PEPTIDE FOR THE SECRETION OF MATURE PROTEINS BY RECOMBINANT YEAST, EXPRESSION CASSETTES, TRANSFORMED YEASTS AND CORRESPONDING PROCESS FOR THE PREPARATION OF PROTEINS		
<b>(54) Titre:</b> APPLICATION DE NOUVEAUX FRAGMENTS D'ADN EN TANT QUE SÉQUENCE CODANT POUR UN PEPTIDE SIGNAL POUR LA SÉCRETION DE PROTÉINES MATURES PAR DES LEVURES RECOMBINANTES, CASSETTES D'EXPRESSION, LEVURES TRANSFORMÉES ET PROCÉDE DE PRÉPARATION DE PROTÉINES CORRESPONDANT		
<b>(57) Abstract</b>  The invention relates to new DNA fragments and to their application as a DNA coding fragment for a signal peptide which can be used for the secretion of proteins, said peptide including a sequence of amino-acids which show a degree of correspondence of at least 60 % with the sequence of amino-acids (I) or (II), preferably with the sequence (II). The sequences (I) and (II) are as follows: (I) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln. (II) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala.		
<b>(57) Abrégé</b>  L'invention concerne de nouveaux fragments d'ADN et leur application à titre de fragment d'ADN codant pour un peptide signal utile pour la sécrétion de protéines, ce peptide comprenant une séquence d'acides aminés qui présente un degré d'homologie d'au moins 60 % avec la séquence d'acides aminés (I) ou (II), de préférence avec la séquence (II). Les séquences (I) et (II) sont comme suit: (I) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln. (II) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala.		

### DESIGNATIONS DE "DE"

Jusqu'à nouvel avis, toute désignation de "DE" dans toute demande internationale dont la date de dépôt international est antérieure au 3 octobre 1990 a effet dans le territoire de la République fédérale d'Allemagne à l'exception du territoire de l'ancienne République démocratique allemande.

### UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AT	Autriche	ES	Espagne	MC	Monaco
AU	Australie	FI	Finlande	MG	Madagascar
BB	Barbade	FR	France	ML	Mali
BE	Belgique	GA	Gabon	MR	Mauritanie
BF	Burkina Faso	GB	Royaume-Uni	MW	Malawi
BG	Bulgarie	GR	Grèce	NL	Pays-Bas
BJ	Bénin	HU	Hongrie	NO	Norvège
BR	Brazil	IT	Italie	RO	Roumanie
CA	Canada	JP	Japon	SD	Soudan
CF	République Centrafricaine	KP	République populaire démocratique de Corée	SE	Suède
CG	Congo	KR	République de Corée	SN	Sénégal
CH	Suisse	LI	Liechtenstein	SU	Union soviétique
CM	Cameroun	LK	Sri Lanka	TD	Tchad
DE	Allemagne, République fédérale d'	LU	Luxembourg	TG	Togo
DK	Danemark			US	Etats-Unis d'Amérique

5. APPLICATION DE NOUVEAUX FRAGMENTS D'ADN EN TANT QUE  
SEQUENCE CODANT POUR UN PEPTIDE SIGNAL POUR LA SECRETION  
DE PROTEINES MATURES PAR DES LEVURES RECOMBINANTES,  
CASSETTES D'EXPRESSION, LEVURES TRANSFORMEES ET PROCEDE DE  
10 PREPARATION DE PROTEINES CORRESPONDANT.

La présente invention a pour objet de nouveaux fragment d'ADN et leur  
utilisation en tant que fragments d'ADN codant pour un peptide signal utile pour  
la sécrétion de protéines hétérologues par des cellules eucaryotes (animales ou  
15 végétales) ou procaryotes (bactéries), plus particulièrement des levures telles que  
des souches de *Saccharomyces*.

Lors de la préparation d'une protéine hétérologue par les techniques de  
l'ADN recombinant, un des objectifs souvent poursuivis est l'obtention d'un  
20 produit qui est sécrété dans le milieu de culture des cellules qui synthétisent la  
protéine hétérologue. En effet, dans le cas notamment de la préparation d'une  
protéine d'intérêt industriel, destinée à être produites en grande quantité, il est  
souhaitable, afin qu'elle conserve les propriétés désirées, que la protéine soit  
produite sous forme mature, c'est à dire dépourvue de tout acide aminé ou  
25 séquence peptidique supplémentaire demeurée fusionnée à la protéine. Par ailleurs,  
il peut être intéressant qu'elle soit sécrétée dans le milieu de culture afin de  
faciliter les opérations de récupération et purification.

Les protéines qui sont sécrétées par une cellule, en particulier par une cellule eucaryote, sont très généralement synthétisées sous forme d'un précurseur polypeptidique comprenant un fragment correspondant à la protéine mature (forme active) et un fragment N-terminal dit fragment "pré", aussi appelé peptide signal, qui intervient dans le mécanisme de sécrétion de la protéine par la cellule. En outre, ce précurseur polypeptidique peut comprendre un ou des fragments additionnels, appelés fragments "pro". Dans ce dernier cas, le précurseur polypeptidique est appelé précurseur "pré-pro" ou premier précurseur. Un fragment "pro" est, dans la majorité des cas, inséré entre le peptide signal et le fragment correspondant à la protéine mature, bien que ceci ne soit pas une règle absolue.

Un peptide signal initie (i) l'insertion de la protéine dans la membrane cellulaire, (ii) la translocation de la protéine au travers de la membrane cellulaire, ou (iii) l'entrée de la protéine dans le réticulum endoplasmique de la cellule en vue de la sécrétion de la protéine par la voie du réticulum endoplasmique. Une fois que le peptide signal a rempli son office, il est normalement détaché par clivage protéolytique pour libérer une protéine mature ou un deuxième précurseur appelé précurseur "pro" qui, tout comme le premier précurseur, n'a pas d'activité biologique, ou qui n'a pas l'activité complète de la protéine mature.

Un fragment "pro" est utile dans la mesure où il bloque ou modifie l'activité de la protéine, permettant ainsi de protéger la cellule contre les effets toxiques éventuels de la protéine ou de protéger la protéine contre d'éventuelles modifications ou dégradations. Il peut aussi intervenir dans une certaine mesure dans le mécanisme de sécrétion. En fin du processus de sécrétion, le fragment "pro" du deuxième précurseur est détaché par clivage protéolytique pour libérer

une protéine mature (forme active).

5 A la jonction du fragment "pré" et du fragment "pro", ainsi que à la jonction du fragment "pro" et du fragment correspondant à la protéine mature, doit se trouver un site de clivage protéolytique qui est reconnu par une des protéases de la cellule dans laquelle la protéine est synthétisée. Ce site de clivage protéolytique est généralement constitué d'une séquence de 2 ou 3 acides aminés ou plus (appelée par la suite séquence de protéolyse) qui est accessible à la protéase sur le précurseur "pro" et qui, si elle existe sur le fragment correspondant à la protéine  
10 mature, n'est pas accessible. Trois cas sont a priori possibles, illustrés ci-dessous en référence à la jonction des fragments "pré" et "pro":

- ou bien la protéase coupe en tête de séquence et, par conséquent, la séquence de protéolyse doit se lire sur le fragment "pro";
- ou bien la protéase coupe en queue de séquence et, par conséquent, la  
15 séquence de protéolyse doit se lire sur le fragment "pré";
- ou bien la protéase coupe en milieu de séquence et, par conséquent, la séquence de protéolyse doit se lire à cheval sur les fragments "pré" et "pro".

20 Ces considérations s'appliquent bien sur de manière similaire au site de clivage placé à la jonction du fragment "pro" et du fragment correspondant à la protéine mature.

25 Pour la construction de précurseurs synthétiques par les méthodes du génie génétique on peut être amené à utiliser des fragments naturels (c'est-à-dire tels que trouvés dans la nature) et à insérer en bonne place un nouveau site de clivage protéolytique ou certains acides aminés de manière à reconstituer un nouveau site

de clivage protéolytique en association avec les fragments, ce nouveau site de clivage étant bien sur reconnu par une des protéases de la cellule dans laquelle le précurseur synthétique est exprimé.

5 Un exemple du type de synthèse et du mode de sécrétion décrit ci-dessus est illustré par le cas de la phéromone sexuelle  $\alpha$  de la levure *S. cerevisiae*, aussi appelé facteur  $\alpha$  (codé à partir du gène MF $\alpha$ 1 ou MF $\alpha$ 2). Le Facteur  $\alpha$  est en effet synthétisé sous forme de précurseur "pré-pro" tel que décrit dans Kurjan et Herskowitz, Cell (1982) 30: 933. La séquence d'acides aminés du fragment "pré"

10 du précurseur du Facteur  $\alpha$  est Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser Ala Leu Ala tandis que celle du fragment "pro" du précurseur du Facteur  $\alpha$  est Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile

15 Ala Ala Lys Glu Glu Gly Val Ser Leu Asp.

De manière surprenante, il a maintenant été trouvé que l'extrémité N-terminale du précurseur d'une enzyme de levure, cette extrémité N-terminale ayant la séquence d'acides aminés :

20 Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala, peut être utilisée à titre de peptide signal pour la sécrétion de protéines hétérologues, ainsi que différents variants de cette extrémité N-terminale. Par "protéine hétérologue" on signifie une protéine qui n'est pas produite naturellement par la cellule hôte ou bien qui est codée par une séquence qui ne

25 provient pas de la cellule hôte.

Conformément à ceci, la présente invention a pour objet un fragment d'ADN

isolé qui code pour un peptide dont la séquence d'acides aminés présente un degré d'homologie d'au moins 60%, de manière préférentielle d'au moins 80%, avec la séquence d'acides aminés (I) ou (II), de préférence avec la séquence (II). Les séquences (I) et (II) sont comme suit:

5 (I) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln.

(II) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala.

10 Par "fragment d'ADN isolé" on signifie un fragment d'ADN dont l'extrémité 3' n'est pas liée par liaison covalente à un fragment d'ADN codant pour une enzyme ayant une activité  $\beta$ -1,3 glucanase telle que en particulier décrite dans Klebl & Tanner, J. Bact, Nov 1989, 171: 6259.

15 Plus particulièrement, un fragment d'ADN selon l'invention code pour un peptide comprenant la séquence d'acides aminés (III) suivante :

(III)  $R_1$ - $R_2$ - $R_3$ -Thr-Thr- $R_4$ -Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-  
                   1                                  5                                  10  
 Phe-Thr-Ala- $R_3$ - $R_4$

20                   15                          19

dans laquelle:

$R_1$  est un acide aminé sélectionné parmi Arg et Lys,

$R_2$  et  $R_4$  sont chacun un acide aminé sélectionné de manière indépendante parmi Ala, Asn, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val,

25  $R_3$  et  $R_5$  sont chacun un acide aminé sélectionné de manière indépendante parmi Asp, Gly, Asn, Pro et Ser, et

$R_4$  est un acide aminé sélectionné parmi Val, Leu, Ala, Cys, Phe, Ile et

Met.

De manière préférée, un fragment d'ADN selon l'invention code pour un peptide comprenant la séquence d'acides aminés (IV) suivante:

5 (IV)  $R_1-R_2-R_3$ -Thr-Thr- $R_4$ -Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala- $R_5$ - $R_6$ - $R_7$

dans laquelle :

$R_1$  est un acide aminé sélectionné parmi Arg et Lys,

10  $R_2$  et  $R_6$  sont chacun un acide aminé sélectionné de manière indépendante parmi Ala, Asn, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val,

$R_3$  et  $R_5$  sont chacun un acide aminé sélectionné de manière indépendante parmi Asp, Gly, Asn, Pro et Ser,

15  $R_4$  est un acide aminé sélectionné parmi Val, Leu, Ala, Cys, Phe, Ile et Met,

$R_7$  est une séquence de protéolyse.

$R_7$  est préférentiellement une séquence de protéolyse  $R_8-R_9-R_{10}$  dans laquelle :

20  $R_8$  est un acide aminé sélectionné parmi Ala, Val, Ser, Cys, Gly, Ile, Leu, Thr,

$R_9$  est un acide aminé sélectionné parmi Ala, Arg, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val et

25  $R_{10}$  est un acide aminé sélectionné parmi Ala, Cys, Gly, Leu, Pro, Gln, Ser et Thr.

Selon l'invention, un fragment d'ADN préféré code pour un peptide comprenant une séquence d'acides aminés sélectionnée parmi les séquences

d'acides aminés (V), (VI), (VII) et (VIII) suivantes :

(V) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln

(VI) Arg-Phe-Ser-Thr-Thr-Val-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln

(VII) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-R<sub>7</sub>

dans laquelle R<sub>7</sub> est tel que défini ci-dessus.

(VIII) Arg-Phe-Ser-Thr-Thr-Val-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-R<sub>7</sub>

dans laquelle R<sub>7</sub> est tel que défini ci-dessus.

De manière tout à fait préférée, R<sub>7</sub> est une séquence dans laquelle R<sub>8</sub> est Val, R<sub>9</sub> est Ser et R<sub>10</sub> est Ala.

Il est tout particulièrement préféré qu'un fragment d'ADN selon l'invention ait pour séquence nucléotidique l'une des séquences (IX), (X), (XI) et (XII) suivantes :

(IX) CGT TTC TCT ACT ACA GTC GCT ACT GCA GCT ACT GCG CTA

TTT TTC ACA GCC TCC CAA,

(X) CGT TTC TCT ACT ACA CTC GCT ACT GCA GCT ACT GCG CTA

TTT TTC ACA GCC TCC CAA,

(XI) CGT TTC TCT ACT ACA GTC GCT ACT GCA GCT ACT GCG CTA

TTT TTC ACA GCC TCC CAA GTT TCA GCT,

(XII) CGT TTC TCT ACT ACA CTC GCT ACT GCA GCT ACT GCG CTA

TTT TTC ACA GCC TCC CAA GTT TCA GCT.

Les peptides codés par les fragment d'ADN selon l'invention comprennent une région hydrophobe possédant une structure en hélice  $\alpha$  comprise entre l'acide aminé en position 3 et l'acide aminé en position 18. Cette structure contribue à les rendre aptes à une utilisation comme peptide signal. Il est connu que la partie hydrophobe des séquences signal est composée en majorité des acides aminés Ala, Cys, Phe, Ile, Leu, Met, Val. On peut donc aussi prévoir que certaines modifications d'acides aminés n'engendrent pas de modification dans l'aptitude du peptide signal à jouer son rôle.

Sous un autre aspect, l'invention propose par conséquent l'application d'un fragment d'ADN selon l'invention à titre de fragment d'ADN codant pour un peptide signal utile pour la sécrétion d'une protéine hétérologue par une cellule hôte dans laquelle la protéine hétérologue est synthétisée. D'une manière générale, l'invention peut être mise en oeuvre dans une cellule procaryote ou eucaryote, de préférence dans cette dernière. La cellule eucaryote peut être par exemple, une cellule de mammifère ou de levure. De manière tout à fait préférée, la sécrétion d'une protéine hétérologue à l'aide d'un peptide signal codé par un fragment selon l'invention est réalisée par une cellule de levure, par exemple du genre *Saccharomyces*, plus particulièrement de l'espèce *S. cerevisiae*.

Bien entendu, en tant que fragments d'ADN codant pour un peptide signal, les fragments d'ADN selon l'invention seront précédés d'un codon d'initiation de la traduction, généralement d'un codon codant pour une méthionine, en particulier un ATG.

Les fragments d'ADN selon l'invention peuvent être utilisés pour la construction d'une cassette d'expression d'une protéine, précédés du codon

d'initiation, seuls ou en combinaison avec d'autres composants, par exemple un fragment "pro". Ces fragments d'ADN peuvent être préparés par synthèse chimique, au moyen d'un synthétiseur d'oligonucléotides par une technique connue de l'homme du métier.

5

L'invention concerne également une cassette d'expression d'une protéine hétérologue comprenant un fragment d'ADN selon l'invention à titre de fragment d'ADN codant pour le peptide signal de la dite protéine hétérologue. De manière détaillée, une cassette d'expression selon l'invention qui comporte donc l'information nécessaire à la sécrétion d'une protéine hétérologue mature comprend, de façon séquentielle, au moins:

10

15

- a) un fragment d'ADN comportant des signaux d'initiation de transcription et de traduction,
- b) un fragment d'ADN selon l'invention,
- c) un fragment d'ADN codant pour une protéine hétérologue mature (avec codon de fin de traduction).

20

Dans une première variante, on peut fusionner directement le fragment b) en phase avec le fragment c), dans la mesure où, à la jonction du fragment b) et du fragment c), il existe une séquence d'ADN codant pour un site de clivage protéolytique de manière à permettre la libération d'une protéine mature en fin du processus d'expression et de sécrétion.

De manière préférée, la séquence d'ADN codant pour un site de clivage protéolytique se lit sur le fragment b).

25

Dans une deuxième variante, une cassette d'expression selon l'invention comprend en outre un fragment d'ADN b') codant pour un fragment peptidique

"pro". Ce fragment b') est fusionné en phase avec le fragment b) et le fragment c), dans la mesure où, à la jonction des fragments b) et b') d'une part et des fragments b') et c) d'autre part, il existe une séquence d'ADN codant pour un site de clivage protéolytique.

5

De nombreux fragments b') peuvent être utilisés dans les cassettes selon l'invention. En particulier, divers fragments b') peuvent être construits de manière synthétique. A titre d'exemple, on indique ci-dessous la construction d'un fragment b') synthétique à partir de la séquence d'ADN codant pour le fragment "pro" du précurseur du Facteur  $\alpha$ . Cette séquence peut être utilisée en totalité ou en partie. Dans un mode de réalisation particulier, on utilise cette séquence déletée de la partie codant pour les acides aminés en position 3 à 42 sur le fragment "pro"; c'est-à-dire la séquence codant pour : Ala Pro Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Asp. Pour former le fragment b'), cette dernière séquence d'ADN est suivie d'une séquence codant pour (i) un peptide comprenant un site de clivage protéolytique ou (ii) un site de clivage protéolytique, ce dernier mode de réalisation étant préféré. Tout particulièrement, ce dernier site de clivage est Lys-Arg ou Arg-Arg, celui-ci étant reconnu par l'endopeptidase de levure yscF qui est codée par le gène KEX2 et qui coupe au niveau de l'extrémité C-terminale du dipeptide Lys-Arg ou Arg-Arg. En résumé, un fragment b') particulier code pour Ala Pro Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Asp Lys Arg.

10

15

20

25

La séquence a) comprend en particulier un promoteur fonctionnel dans la cellule dans laquelle on souhaite synthétiser la protéine hétérologue codée par le fragment c), de préférence un promoteur fonctionnel chez la levure. On peut citer par exemple des promoteurs constitutifs de levure dont la fonctionnalité a été

confirmée par la transcription de gènes codant pour des protéines hétérologues codant pour des protéines hétérologues tels que les promoteurs PGK, ENO1, MF $\alpha$ 1, ou encore des promoteurs inductibles, tels que PH05, GAL1. Par exemple, lorsqu'on utilise dans la cassette d'expression des éléments du gène MF $\alpha$ 1 de levure, on pourra utiliser le promoteur du gène MF $\alpha$ 1.

5

Enfin, les cassettes d'expression peuvent aussi comprendre un fragment d'ADN d) comportant des signaux de terminaison de la transcription, de préférence fonctionnels chez la levure, par exemple celui du gène PGK.

10

De façon générale, une cassette d'expression selon l'invention peut être introduite dans une cellule procaryote ou eucaryote, de préférence eucaryote telle que cellule de mammifère ou de levure; une cellule de levure étant tout particulièrement préférée. Cette introduction peut être réalisée en plaçant la cassette dans un plasmide à répllication autonome ou dans une construction destinée à l'intégration pour être introduite directement dans le génome de la levure, de manière à obtenir dans les deux cas une cellule transformée.

15

Lorsque le plasmide est autonome, il comportera des éléments assurant sa partiton et sa répllication; par exemple, une origine de répllication telle que celle du plasmide 2 $\mu$  de levure. En outre, le plasmide pourra comporter des éléments de sélection tels que le gène URA3 ou LEU2 qui assurent la complémentation de levure ura3 ou leu2. En particulier, on pourra avantageusement utiliser le gène URA3 délété de son promoteur (URA3-d).

20

25

Ces plasmides peuvent également comporter des éléments assurant leur répllication dans les bactéries, lorsque le plasmide doit être un plasmide navette,

par exemple une origine de répllication telle que celle du pBR322, un gène marqueur de sélection tel que Amp<sup>R</sup> et/ou d'autres éléments connus de l'homme du métier.

5           En accord avec ce qui précède, la présente invention concerne également une cellule transformée par (i) un fragment d'ADN selon l'invention ou (ii) une cassette d'expression selon l'invention, soit insérée dans un plasmide, soit intégrée dans le génome de la cellule.

10           Lorsque le promoteur est celui du gène MF $\alpha$ 1, la cellule de levure transformée est de préférence de type sexuel MAT $\alpha$ . On utilisera, par exemple, une souche de génotype ura3 ou leu2 ou autre, complémentée par le plasmide pour assurer le maintien du plasmide dans la levure par une pression de sélection appropriée.

15           Enfin, l'invention a pour objet un procédé de préparation d'une protéine hétérologue, caractérisé en ce que l'on cultive une cellule selon l'invention et, en ce que l'on récupère la dite protéine dans le milieu de culture. Ce procédé s'applique à la préparation de toute protéine de nature hétérologue. Parmi ces protéines, on peut notamment citer l'hirudine ou des défensines, par exemple la  
20           défensine A.

          Plus particulièrement, l'invention concerne un procédé de sécrétion de l'hirudine sous forme mature à partir des souches de levures transformées selon l'invention.

25

          L'invention s'applique tout particulièrement bien à la production d'hirudine, c'est pourquoi un des exemples illustratifs de l'invention concerne cette

protéine. En effet l'hirudine, dont la source principale se trouve dans les glandes salivaires des sangsues médicinales est un inhibiteur très spécifique et très efficace de la thrombine. Il s'agit donc d'un agent thérapeutique très intéressant dont l'utilisation en clinique exige une très grande pureté du produit, et qui est donc un candidat intéressant à la production par génie génétique.

Un certain nombre de variants naturels de l'hirudine ont été identifiés et désignés par HV1, HV2, HV3. Par la suite ces variants naturels ainsi que d'autres analogues ont été préparés par génie génétique dans diverses cellules hôtes, comme cela est décrit par exemple dans les publications européennes de brevet EP-A-0200655, EP-A-0273800 au nom de la Demanderesse. La comparaison de l'hirudine synthétisée par *Escherichia coli* (*E. coli*) et par une levure du genre *S. cerevisiae* a montré que l'hirudine synthétisée par *E. coli* reste intracellulaire et doit donc être purifiée à partir d'un très grand nombre de polypeptides de *E. coli*. Il est donc particulièrement intéressant de pouvoir faire exprimer un gène de l'hirudine dans la levure de façon à obtenir une hirudine sécrétée sous forme mature, et sans que la levure ne produise de substances pyrogènes ou toxiques pour l'homme.

L'invention s'applique donc à toutes les molécules d'hirudines, c'est à dire, variants naturels de l'hirudine tel quels ou ayant subi une ou plusieurs mutations tout en conservant leur activité antithrombotique; ce dernier type de variant étant appelé analogue. Les exemples ci-après concerneront plus particulièrement l'analogue désigné par rHV2Lys47 (pour recombinant variant HV2 ayant subi une mutation de l'acide aminé Asp en position 47 en acide aminé Lys), décrit dans la publication de brevet EP-A-02738000 déjà mentionnée.

L'invention s'applique également à la production de défensines. Les défensines, aussi appelées phormicines, sont des peptides originellement extraits de l'hémolymph des certains insectes, les Diptères, qui ont une activité bactéricide sur les germes Gram-positifs. Ces défensines sont plus amplement  
5 décrites dans la demande de brevet européenne EP-A- 349 451. La défensine A est un peptide basique ayant pour séquence Ala Thr Cys Asp Leu Leu Ser Gly Thr Gly Ile Asn His Ser Ala Cys Ala Ala His Cys Leu Leu Arg Gly Asn Arg Gly Gly Tyr Cys Asn Gly Lys Gly Val Cys Val Cys Arg Asn. La défensine B ne diffère de la défensine A que par l'acide aminé en position 32 où une arginine remplace  
10 une glycine.

- Les exemples ci-après permettront de mettre en évidence d'autres caractéristiques et avantages de la présente invention. Ces exemples seront illustrés par les figures suivantes :
- 15 - la figure 1 représente la structure schématique du plasmide pTG2958.
  - la figure 2 représente la structure schématique des vecteurs M13TG3839 et M13TG3841. Pour M13TG3839 les cadres hachurés aux extrémités correspondent à M13TG103 et pour M13TG3841 à M13TG3149.
  - la figure 3 représente la structure schématique du vecteur M13TG3845. Le  
20 cadre hachuré correspond à M13TG3149.
  - la figure 4 représente la structure schématique du plasmide pTG3828.
  - la figure 5 représente la structure schématique du plasmide pTG3864.

EXEMPLE 1 : Construction des vecteurs d'expression de l'hirudine :  
pTG3864, pTG3867, pTG3894 et pTG3884.

A. Construction du vecteur M13TG3845

5 Le plasmide pTG2958 (figure 1) est peu différent du plasmide pTG1833 décrit dans la publication européenne de brevet EP-A-252854 porteur de la séquence codante pour rHV2Asp47. Le plasmide pTG2958 ne contient pas le site de restriction HindIII artificiellement introduit. Le plasmide pTG2958 contient :

- 10 - un fragment de 1217 paires de bases correspondant à la région 5' du gène MF $\alpha$ 1 (contenant le promoteur, la séquence codant pour le peptide signal, la région "pro" et une séquence codant pour le peptide Lys-Arg), et 4 paires de bases (site BglII<sup>o</sup>, traitement par Klenow),
- un fragment de 234 paires de bases contenant l'ADN complémentaire de rHV2Lys47,
- 15 - un fragment de 243 paires de bases comprenant le terminateur PGK de levure,
- le fragment PvuII-EcoRI de pBR322 comprenant entre autres l'origine de réplication de ce plasmide et le gène de résistance à l'ampicilline (2292 paires de bases),
- le fragment EcoRI-HindIII du plasmide 2 $\mu$  de la levure (forme B), contenant
- 20 le gène LEU2 de levure, sous forme délétée et inséré dans le site PstI,
- un fragment HindIII-SmaI du gène URA3 de levure.

25 Le fragment NcoI-NcoI du vecteur pTG2958 qui porte les séquences LEU2-d, 2 $\mu$  et URA3 est remplacé par le fragment NcoI-NcoI de pTG2800 décrit dans la publication européenne de brevet EP-A-0268501 qui porte les séquences du plasmide 2 $\mu$  et du gène URA3 délété de son promoteur (URA3-d) pour donner pTG2877.

Le vecteur M13TG3839 (figure 2) dérive de M13TG103 [Kieny, M.P. et

*al.* (1983) Gene 26, 91-99] dans lequel le fragment HindIII-HindIII de pTG2877 est introduit dans le même site. Un site de restriction SalI est introduit dans ce vecteur en aval du codon de terminaison de traduction de la région codant pour rHV2Lys47 par mutagenèse dirigée à l'aide de l'oligonucléotide suivant :

5 5' CAATGAAAAATGGTCGACTATCAATCATAG pour donner M13TG3839  
SalI. Un site de restriction SphI est alors introduit en amont de la cassette d'expression en éliminant la séquence URA3-d par mutagenèse dirigée à l'aide de l'oligonucléotide suivant :  
5' GACGGCCAGTGAATTGGCATGCTATTGATAAGATTAAAG pour donner  
10 M13TG3840.

Le vecteur M13TG131 [Kieny M.P. *et al.* (1983) Gene 26, 91-99] est clivé par PstI, les extrémités rendues franche par traitement à l'aide du fragment Klenow de l'ADN polymérase I pour être ensuite reliqué sur lui même pour donner M13TG3160. Ce vecteur est ensuite clivé par SmaI et EcoRV puis reliqué  
15 pour donner M13TG3149.

Le fragment SphI-SalI de M13TG3840 (décrit ci-dessus) portant la cassette d'expression de rHV2Lys47 (sans séquence terminatrice de transcription) est introduit dans le site SphI-SalI de M13TG3149 pour donner M13TG3841 (figure 2).

20 La séquence codant pour les acides aminés en position 3 à 42 sur le fragment "pro" du précurseur de MF $\alpha$ 1 est éliminée de M13TG3149 par mutagenèse dirigée et un site de restriction SmaI introduit en utilisant l'oligonucléotide suivant :

5' CTCGCGATTAGCTGCTCCCGGGTTATGTTTATAAAT, pour donner ce que  
25 l'on appelle par la suite une séquence "pro" délétée.

On obtient ainsi M13TG3842. Un site de restriction BamHI détruisant l'ATG du précurseur du facteur  $\alpha$  est introduit dans ce vecteur par mutagenèse dirigée avec

l'oligonucléotide suivant :

5' AATATAAACGATTAAAAGGATCCGATTTCTTCAATTTT

On obtient alors M13TG3843. Après phosphorylation, les oligonucléotides suivants:

5' GATCCGTTTCTCTACTACAGTCGCTACTGCAGCTACTGCGCTATT  
GCAAAGTGATGATGTCAGCGATG 5'

et

5' TTTCACAGCCTCCCAAGTTTCAGCTGCTCCC  
ACGTCGATGACGCGATAAAAAGTGTGGAGGGTCAAAGTCGACGAGGG 5'  
sont insérés dans le vecteur M13TG3843 coupé par BamHI et SmaI introduisant  
ainsi la séquence XI donc sans ATG. De manière à restaurer l'ATG, le site BamHI  
est éliminé par mutagenèse dirigée en utilisant l'oligonucléotide suivant :

5' AATATAAACGATTAAAAGAATGCGTTTCTCTACTACAGTC

pour donner le vecteur M13TG3845 (figure 3) qui comporte :

- le promoteur du gène MF $\alpha$ 1, suivi d'un codon ATG, en tant que fragment a)
- la séquence XI en tant que fragment b),
- la séquence "pro" délétée du gène MF $\alpha$ 1, suivi des codons codant pour Lys-Arg en tant que fragment b'),
- la séquence codant pour rHV2Lys47 en tant que fragment c),
- une partie du vecteur M13TG3149.

#### B. Construction du plasmide pTG3864

Le plasmide pTG848 décrit dans la publication européenne de brevet EP-A-0252854 est digéré par BglII puis reliqué pour donner pTG2886. Le grand fragment HindIII-EcoRI de pTG2886 est ligué en présence de ligase T4 au fragment HindIII-EcoRI de 2,1 kb du plasmide pFL1 [Parent, S.A. *et al.* (1985) Yeast 1, 83-138] qui porte la séquence du plasmide 2 $\mu$  de *S. cerevisiae* pour donner le plasmide pTG2886 LEU2-d, URA3-d. Le fragment HindIII de 0,9 kb du

- plasmide pTG2800 décrit dans la publication européenne de brevet EP-A-0258501 portant le gène URA3-d est alors inséré dans le site HindIII de ce plasmide pour donner pTG2886 URA3-d, delta LEU2-d. Le fragment SmaI-BglII de M13TG131 [Kieny *et al.* (1983) Gene 26, 91-99] qui possède plusieurs sites de restriction est ensuite introduit dans ce plasmide pour donner pTG3828 (figure 4) qui comporte:
- la séquence du gène URA3 déléetée de son promoteur (URA3-d),
  - des sites de restriction venant de M13TG131 permettant l'insertion des éléments de l'expression d'un gène hétérologue, rHV2Lys47 dans le cas présent,
  - le terminateur de transcription du gène PGK de levure
  - un fragment de pBR322 qui permet la réplication et la sélection chez *E. coli*,
  - un fragment du plasmide 2 $\mu$  qui possède les éléments structuraux nécessaires à la réplication et à l'équipartition mitotique dans la levure.

Le fragment SphI-SalI du vecteur M13TG3845 (figure 3) est introduit dans le plasmide pTG3828 digéré par SphI et SalI pour donner le vecteur d'expression pTG3864 (figure 5).

#### C. Construction du vecteur d'expression pTG3867

Pour supprimer complètement la séquence codant pour le précurseur du gène MF $\alpha$ 1 on effectue une mutagénèse dirigée sur M13TG3845 (figure 3) à l'aide de l'oligonucléotide suivant :

5' GCCTCCCAAGTTTCAGCTATTACGTATACAGACTGC

pour obtenir le vecteur M13TG3846. Le fragment SphI-SalI de ce vecteur est introduit dans le plasmide pTG3828 (figure 4) digéré par SphI et SalI pour donner le vecteur d'expression pTG3867. Dans ce vecteur, la séquence XI et celle codant pour rHV2Lys47 sont adjacentes. Pour obtenir la structure schématique de ce plasmide, il suffit sur la figure 5 de retirer la séquence "pro" déléetée de MF $\alpha$ 1.

D. Construction du vecteur d'expression pTG3894

Un site SmaI est créé dans la séquence codant pour le fragment "pro" du précurseur du facteur  $\alpha$  par mutagenèse dirigée sur M13TG3841 (figure 2) grâce à l'oligonucléotide suivant :

5' TCCGCATTAGCTGCTCCCGGGAACACTACAACAGAA

pour obtenir M13TG3869. Ce vecteur est ensuite digéré par SphI et SmaI, le petit fragment est isolé et il est ligué au grand fragment SphI et SmaI de M13TG3845 (figure 3) pour donner le vecteur M13TG3891. Le fragment SphI-SalI de ce vecteur est introduit dans le plasmide pTG3828 (figure 4) ouvert aux sites SphI et SalI pour donner le vecteur d'expression pTG3894 qui comporte donc la séquence "pro" du précurseur du facteur  $\alpha$  mutée. Pour obtenir la structure schématique de ce plasmide, il suffit sur la figure 5 de remplacer la séquence "pro" déléetée de MF $\alpha$ 1 par la séquence "pro" mutée.

15 E. Construction du vecteur d'expression pTG3884

La séquence XI est modifiée par mutagenèse dirigée sur M13TG3845 en utilisant l'oligonucléotide suivant :

5' GTTTCCTCTACTACTCGCTACTGC

La modification d'une seule base donne la séquence XII et induit le remplacement d'une valine par une leucine en tant qu'acide aminé R<sub>4</sub> du peptide signal. On obtient ainsi le bactériophage M13TG3846. Le fragment SphI-SalI de M13TG3846 est introduit dans le plasmide pTG3828 (figure 4) ouvert aux sites SphI et SalI pour donner le vecteur d'expression pTG3884 qui comporte :

- la séquence du gène URA3 déléetée de son promoteur (URA3-d),
- 25 - le promoteur du gène MF $\alpha$ 1, suivi d'un codon ATG, en tant que fragment a),
- la séquence XII en tant que fragment b),
- la séquence "pro" déléetée du gène MF $\alpha$ 1, suivi des codons codant pour Lys-

Arg, en tant que fragment b'),

- la séquence codant pour rHV2Lys47 en tant que fragment c),
- le terminateur du gène codant pour PGK de la levure.
- un fragment de pBR322,
- 5 - un fragment du plasmide 2 $\mu$ .

EXEMPLE 2 :        Production de rHV2Lys47 dans le surnageant de culture en fonction du plasmide utilisé.

10        Une souche de levure de l'espèce *Saccharomyces cerevisiae* de génotype MAT $\alpha$ , ura3-251,-373,-328, leu2-3,-112, his3, pep4-3 est transformée par les plasmides pTG3864, pTG3867, pTG3894 et pTG3884 par la méthode de l'acétate de lithium [Ito, H. et al. J. Bactériol. (1983) 153: 163] et les prototrophes Ura+ sont sélectionnés. Ils sont ensuite mis en culture en erlenmeyer à 30°C sur un

15        milieu sélectif (0,7% de bases azotées pour levures (Yeast Nitrogen Base), 0,5% de casamino acides et 1% de glucose). Après 48 heures de culture, on sépare cellules et surnageant par centrifugation et l'activité inhibitrice de la thrombine est déterminée dans le surnageant en utilisant le test colorimétrique (activité

20        protéolytique sur un substrat synthétique, le chromozyme TH - Boehringer Mannheim). Le tableau I présente les résultats des dosages; chaque valeur correspond à la moyenne de deux expériences indépendantes. L'activité de rHV2Lys47 est exprimée en ATU/ml de surnageant.

Tableau I

Plasmide	ATU/ml
pTG3894	40
pTG3864	50
pTG3867	130
pTG3884	125

5 Dans tous les cas on mesure une activité anti-thrombine. La protéine rHV2Lys47 produite par la levure est donc excrétée dans le surnageant. De plus elle est sécrétée sous forme active. Les meilleurs résultats sont obtenus pour les souches transformées par pTG3884 et pTG3867.

10 Le contenu en protéine des surnageants est analysé par HPLC. Le pic majeur obtenu correspond bien à celui de rHV2Lys47 (sous sa forme à 65 acides aminés) et la détermination de la séquence N-terminale confirme l'obtention d'une molécule correctement synthétisée.

15 EXEMPLE 3 : Construction d'un vecteur d'expression de la défensine A d'insectes : pTG4826.

20 A: Synthèse d'une séquence d'ADN codant pour la défensine A d'insecte.

La synthèse se fait en deux blocs assemblés grâce à leurs extrémités cohésives KpnI. Le premier bloc comprend 3 oligonucléotides numérotés de 1 à 3 et le second bloc, 6 oligonucléotides numérotés de 4 à 9. Leur séquence et la position des oligonucléotides (dernière ligne du tableau; les ronds représentent la partie 5' de l'oligonucléotide) sont donnés dans le tableau II.

TABLEAU II

n°	Séquence
1	5' AGCTTGGACAAGAGAGCTACCTGTGACTTGTGTCGGGTAC
2	5' GGTAGCTCTCTTGTOCA
3	5' CGGACAACAAGTCACA
4	5' CGGTATTAACCACTCCGCTTGCTGCTCACTGTTTGTG
5	5' AGCACAAGCGGAGTGGTTAATACCGGTAC
6	5' AGAGGTAACAGAGGTGGCTACTGTAACGGTAAGGGTGT
7	5' AGTAGCCACCTCTGTTACCTCTCAACAAACAGTGAGC
8	5' TTGTGTTTGTAGAACTAAGGATCCG
9	5'AATTCGGATCCTTAGTTTCTACAAACACAAACACCCTTACCGT TAC

La séquence obtenue est la suivante :

HindIII

+1

KpnI

10

20

AGC TTG GAC AAG AGA GCT ACC TGT GAC TTG TTG TCC GGT ACC  
Ala Thr Cys Asp Leu Leu Ser Gly Thr

20

25

GGT ATT AAC CAC TCC GCT TGT GCT GCT CAC TGT TTG TTG AGA  
Gly Ile Asn His Ser Ala Cys Ala Ala His Cys Leu Leu Arg

23

30

GGT AAC AGA GGT GGC TAC TGT AAC GGT AAG GGT GTT TGT GTT  
 Gly Asn Arg Gly Gly Tyr Cys Asn Gly Lys Gly Val Cys Val

5                   40                   BamHI-EcoRI  
 TGT AGA AAC TAA GGATCCG  
 Cys Arg Asn

10       La synthèse du premier bloc utilise les oligonucléotides 4, 5, 6, 7, 8 et 9 et s'effectue de la façon suivante :

- les oligonucléotides 5, 6, 7 et 8 sont tout d'abord phosphorylés à leur extrémités 5' pour éviter la formation de polymères au cours de l'assemblage. Pour chacun de ces oligonucléotides, 100 picomoles sont traitées à la polynucléotide kinase, 2 unités dans un volume final de 20 µl de Tris HCl 60 µM à pH 7,5; 10 µM de MgCl<sub>2</sub>; 8 µM de dithiothréitol (tampon de kination)
- 15       contenant 3,3 picomoles d'ATPy marqué avec <sup>32</sup>P (5000 Ci/mmmole). Après 15 minutes d'incubation à 37°C, 5 µmoles d'ATP non marqué sont ajoutées.
- après incubation à 37°C pendant 30 min. 75 picomoles des oligonucléotides 5, 6, 7 et 8 sont mélangés, chauffés à 95°C pendant 3 min., puis les
- 20       oligonucléotides 4 et 9 sont ajoutés dans un volume final de 90 µmoles de tampon de kination décrit ci-dessus. L'ensemble est chauffé à 95°C pendant 3 min. puis refroidi lentement en 2 heures à 37°C.
- 25 picomoles de ces oligonucléotides hybridés sont soumis au traitement par la
- ligase T4 pendant une heure à 15°C. Ce mélange réactionnel (1 picomole) est
- 25       ensuite ajouté à 50 ng du du bactériophage M13TG131 [Kieny M.P. *et al.* (1983) *Gene* 26, 91-99] traité par EcoRI et KpnI (1 heure à 15°C). Le mélange de ligation est utilisé pour transformer les cellules compétentes de la souche *E. coli* JM103 [Messing J. *et al.* (1981), *Nucleic Acid Res.* 9, 309]. Un clone présentant la séquence recherchée est isolé, il est appelé M13TG3821.

30       La synthèse du second bloc utilise les oligonucléotides 1, 2 et 3 et s'effectue selon la même procédure que celle décrite pour la synthèse du premier

bloc. Dans ce cas, seul l'oligonucléotide 2 est phosphorylé à son extrémité 5'.

Ce second bloc est cloné entre les sites HindIII et KpnI du bactériophage M13TG3821 et un clone portant la séquence d'ADN codant pour la défensine A (figure 5) est isolé; il est appelé M13TG3849.

5

B. Construction du plasmide d'expression de la défensine A: pTG4839.

Le fragment SphI - SmaI de 1045 paires de bases du bactériophage M13TG3846 décrit précédemment (Exemple 1, E.) est transféré dans le vecteur M13TG3869 décrit précédemment (Exemple 1, D.) préalablement digéré par SphI et SmaI. On obtient ainsi le vecteur M13TG4803 qui porte :

10

- le promoteur du gène MF $\alpha$ 1, suivi d'un codon ATG,
- la séquence XII,
- la séquence "pro" mutée du gène MF $\alpha$ 1 suivi des codons codant pour Lys-Arg,
- la séquence codant pour rHV2Lys47.

15

Afin de remplacer la séquence codant pour rHV2Lys47 par celle codant pour la défensine A d'insecte on introduit un site HindIII dans la séquence codant pour "pro" mutée de MF $\alpha$ 1 à l'aide de l'oligonucléotide de séquence :

5' GAAGGGGTAAGCTTGGATAAA

20

Puis on introduit le fragment HindIII - BamHI de M13TG3849 décrit précédemment (Exemple 3, A.) qui porte la séquence synthétique codant pour la défensine A dans ce vecteur préalablement traité par HindIII et BamHI pour éliminer la séquence codant pour rHV2Lys47.

25

Le fragment SphI-SalI de M13TG4813 est introduit dans le plasmide pTG3828 (figure 4) ouvert aux sites SphI et SalI pour donner le vecteur d'expression pTG4839 qui comporte :

- la séquence du gène URA3 déléetée de son promoteur (URA3-d),
- le promoteur du gène MF $\alpha$ 1, suivi d'un ATG,

- la séquence XII,
  - la séquence "pro" mutée du gène MF $\alpha$ 1 suivi des codons codant pour Lys-Arg,
  - la séquence synthétique codant pour la défensine A d'insectes,
  - le terminateur du gène codant pour PGK de la levure.
- 5
- un fragment de pBR322,
  - un fragment du plasmide 2 $\mu$ .

EXEMPLE 4 : Production de défensine A dans le surnageant de culture.

10 Une souche de levure de l'espèce *Saccharomyces cerevisiae* de génotype MAT $\alpha$ , ura3-251,-373,-328, leu2-3,-112, his3, pep4-3 est transformée par le plasmide pTG4839 par la méthode de l'acétate de lithium [Ito, H. *et al.* J. Bactériol. (1983) 153: 163] et les prototrophes Ura<sup>+</sup> sont sélectionnés. Ils sont ensuite mis en culture en erlenmeyer à 30°C sur un milieu sélectif (0,7% de bases

15 azotées pour levures (Yeast Nitrogen Base), 0,5% de casamino acides et 1% de glucose). Après 48 heures de culture, on sépare cellules et surnageant par centrifugation et le surnageant est filtré sur un filtre de 22  $\mu$  puis passé sur cartouche Sep-Pak C18. Le matériel fixé est élué avec 60% d'acétonitrile, 0,1% d'acide trifluoro acétique dans de l'eau et séché sous vide. L'activité

20 antibactérienne de la défensine A est ensuite mise en évidence par un test d'étalement sur agar ou sur gélose ensemencé de germes bactériens (*Micrococcus luteus*) conformément à la procédure décrite par Lambert *et al.* (1989) PNAS 86: 262-266.

25 Dans le surnageant des levures transformées par le plasmide pTG4839 on détecte effectivement une activité antibactérienne. La protéine défensine A produite par la levure est donc excrétée dans le surnageant. De plus elle est sécrétée sous forme active.

Le contenu en protéine des surnageants est analysé par HPLC. Le pic majeur obtenu correspond bien à celui de la défensine A d'insecte et la détermination de la séquence de la protéine confirme l'obtention d'une molécule correctement synthétisée.

## REVENDEICATIONS

1. Un fragment d'ADN isolé qui code pour un peptide dont la séquence d'acides aminés présente un degré d'homologie d'au moins 60% avec la séquence d'acides aminés de formule (I) ou (II)
- (I) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln et
- (II) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala.
2. Un fragment d'ADN selon la revendication 1 qui code pour un peptide dont la séquence d'acides aminés présente un degré d'homologie d'au moins 80% avec la séquence d'acides aminés (I) ou (II) de formule
- (I) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln et
- (II) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-Val-Ser-Ala.
3. Un fragment d'ADN selon la revendication 1 ou 2 qui code pour un peptide comprenant la séquence d'acides aminés (III)
- (III)  $R_1$ - $R_2$ - $R_3$ -Thr-Thr- $R_4$ -Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-  
                   1                                  5                                  10  
 Phe-Thr-Ala- $R_5$ - $R_6$   
                   15                                  19
- dans laquelle :
- $R_1$  est un acide aminé sélectionné parmi Arg et Lys,  
 $R_2$  et  $R_4$  sont chacun un acide aminé sélectionné de manière indépendante parmi Ala, Asn, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val

$R_3$  et  $R_5$  sont chacun un acide aminé sélectionné de manière indépendante parmi Asp, Gly, Asn, Pro et Ser, et

$R_4$  est un acide aminé sélectionné parmi Val, Leu, Ala, Cys, Phe, Ile et Met.

- 5 4. Un fragment d'ADN selon la revendication 1 ou 2 qui code pour un peptide comprenant la séquence d'acides aminés (IV)

(IV)  $R_1$ - $R_2$ - $R_3$ -Thr-Thr- $R_4$ -Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala- $R_5$ - $R_6$ - $R_7$

dans laquelle:

- 10  $R_1$  est un acide aminé sélectionné parmi Arg et Lys,  
 $R_2$  et  $R_6$  sont chacun un acide aminé sélectionné de manière indépendante parmi Ala, Asn, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val  
 $R_3$  et  $R_5$  sont chacun un acide aminé sélectionné de manière indépendante parmi Asp, Gly, Asn, Pro et Ser,  
15  $R_4$  est un acide aminé sélectionné parmi Val, Leu, Ala, Cys, Phe, Ile et Met, et  
 $R_7$  est une séquence de protéolyse.

5. Un fragment d'ADN selon la revendication 4 dans lequel  $R_7$  est une séquence de protéolyse  $R_8$ - $R_9$ - $R_{10}$  dans laquelle:

- 20  $R_8$  est un acide aminé sélectionné parmi Ala, Val, Ser, Cys, Gly, Ile, Leu, Thr,  
 $R_9$  est un acide aminé sélectionné parmi Ala, Arg, Cys, Gln, Gly, His, Ile, Leu, Met, Phe, Pro, Ser, Thr, Trp, Tyr et Val et  
 $R_{10}$  est un acide aminé sélectionné parmi Ala, Cys, Gly, Leu, Pro, Gln, Ser et Thr.

25

6. Un fragment d'ADN selon la revendication 1 ou 2 qui code pour un peptide comprenant une séquence d'acides aminés sélectionnée parmi les

séquences d'acides aminés (V) et (VI)

(V) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln;

(VI) Arg-Phe-Ser-Thr-Thr-Val-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln.

7. Un fragment d'ADN selon l'une des revendications 1, 2, 4 et 5 qui code pour un peptide comprenant une séquence d'acides aminés sélectionnée parmi les séquences d'acides aminés de formule (VII) et (VIII)

(VII) Arg-Phe-Ser-Thr-Thr-Leu-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-R<sub>1</sub>,

dans laquelle R<sub>1</sub> est une séquence de protéolyse;

(VIII) Arg-Phe-Ser-Thr-Thr-Val-Ala-Thr-Ala-Ala-Thr-Ala-Leu-Phe-Phe-Thr-Ala-Ser-Gln-R<sub>2</sub>,

dans laquelle R<sub>2</sub> est une séquence de protéolyse.

8. Un fragment d'ADN selon la revendication 7 qui code pour un peptide comprenant une séquence d'acides aminés sélectionnée parmi les séquences d'acides aminés de formule (VII) et (VIII) dans lesquelles R<sub>1</sub> est Val-Ser-Ala.

9. Application d'un fragment d'ADN selon l'une des revendications 1 à 8, à titre de fragment d'ADN codant pour un peptide signal utile pour la sécrétion d'une protéine hétérologue par une cellule dans laquelle la protéine hétérologue est synthétisée.

10. Application d'un fragment d'ADN selon la revendication 9, caractérisé en ce que le fragment d'ADN code pour un peptide signal utile pour la sécrétion

d'une protéine hétérologue par une cellule de levure dans laquelle la protéine hétérologue est synthétisée.

- 5 11. Une cassette d'expression d'une protéine hétérologue comprenant au moins:
- a) un fragment d'ADN comportant des signaux d'initiation de transcription et de traduction,
  - b) un fragment d'ADN selon l'une des revendications 1 à 8 et,
  - c) un fragment d'ADN codant pour la protéine hétérologue mature.
- 10 12. Une cassette selon la revendication 11 comprenant en outre un fragment d'ADN b') codant pour un fragment peptidique "pro".
- 15 13. Une cassette selon la revendication 12 comprenant un fragment d'ADN b') codant pour un fragment peptidique "pro" ayant pour séquence Ala Pro Gly Leu Leu Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Asp Lys Arg.
- 20 14. Une cassette selon l'une des revendications 11 à 13, caractérisée en ce le fragment a) comporte un promoteur fonctionnel dans une cellule de levure et un codon d'initiation de traduction ATG.
- 25 15. Une cassette d'expression selon l'une des revendications 11 à 14, caractérisée en ce que le fragment d'ADN c) code pour une hirudine.
16. Une cassette d'expression selon la revendication 15, caractérisée en ce que le fragment d'ADN c) code pour le variant hirudine rHV2Lys47.

17. Une cassette d'expression selon l'une des revendications 11 à 14, caractérisée en ce que le fragment d'ADN c) code pour une défensine d'insectes.
18. Une cassette d'expression selon la revendication 17, caractérisée en ce que  
5 le fragment d'ADN c) code pour la défensine A.
19. Un vecteur plasmidique comprenant une cassette d'expression selon l'une des revendications 11 à 18.
- 10 20. Un vecteur plasmidique selon la revendication 19, caractérisé en ce qu'il comporte en fragment du plasmide 2 $\mu$  de levure.
21. Un vecteur plasmidique selon la revendication 19 ou 20, caractérisé en ce qu'il comporte, en tant que gène de sélection, le gène URA3 délété de son  
15 promoteur.
22. Une cellule transformée par un vecteur plasmidique selon l'une des revendications 19 à 21 ou ayant intégré dans son génome une cassette d'expression selon l'une des revendications 11 à 18.  
20
23. Une cellule de levure selon la revendication 22.
24. Un procédé de préparation d'une protéine hétérologue, caractérisé en ce que l'on cultive une cellule selon la revendication 22 ou 23 et, en ce que l'on récupère  
25 la dite protéine dans le milieu de culture.
25. Un procédé selon la revendication 24, caractérisé en ce que la cellule est

une cellule de levure.

26. Un procédé selon la revendication 24 ou 25, caractérisé en ce que ladite protéine est une hirudine.

5

27. Un procédé selon la revendication 24 ou 25, caractérisé en ce que ladite protéine est une défensine d'insectes.

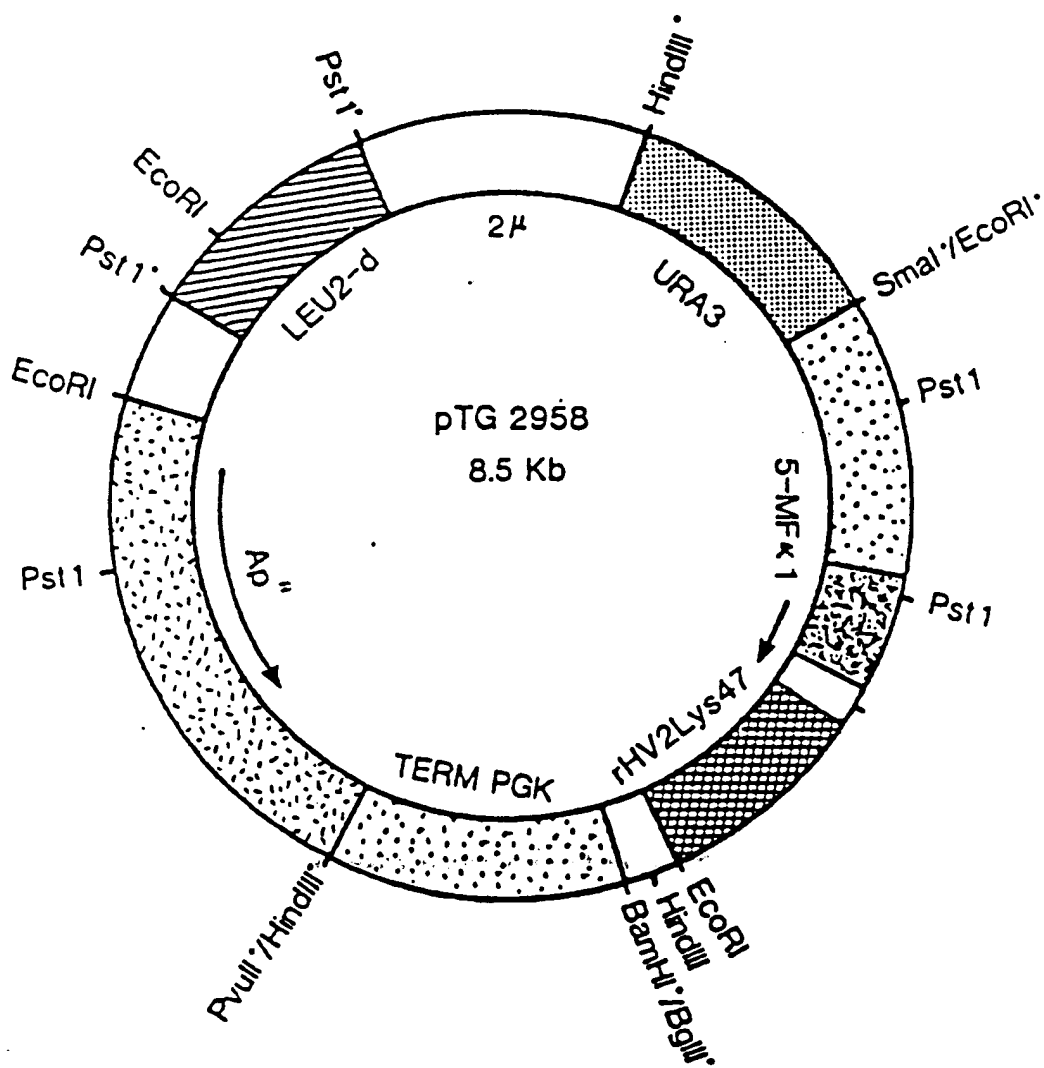
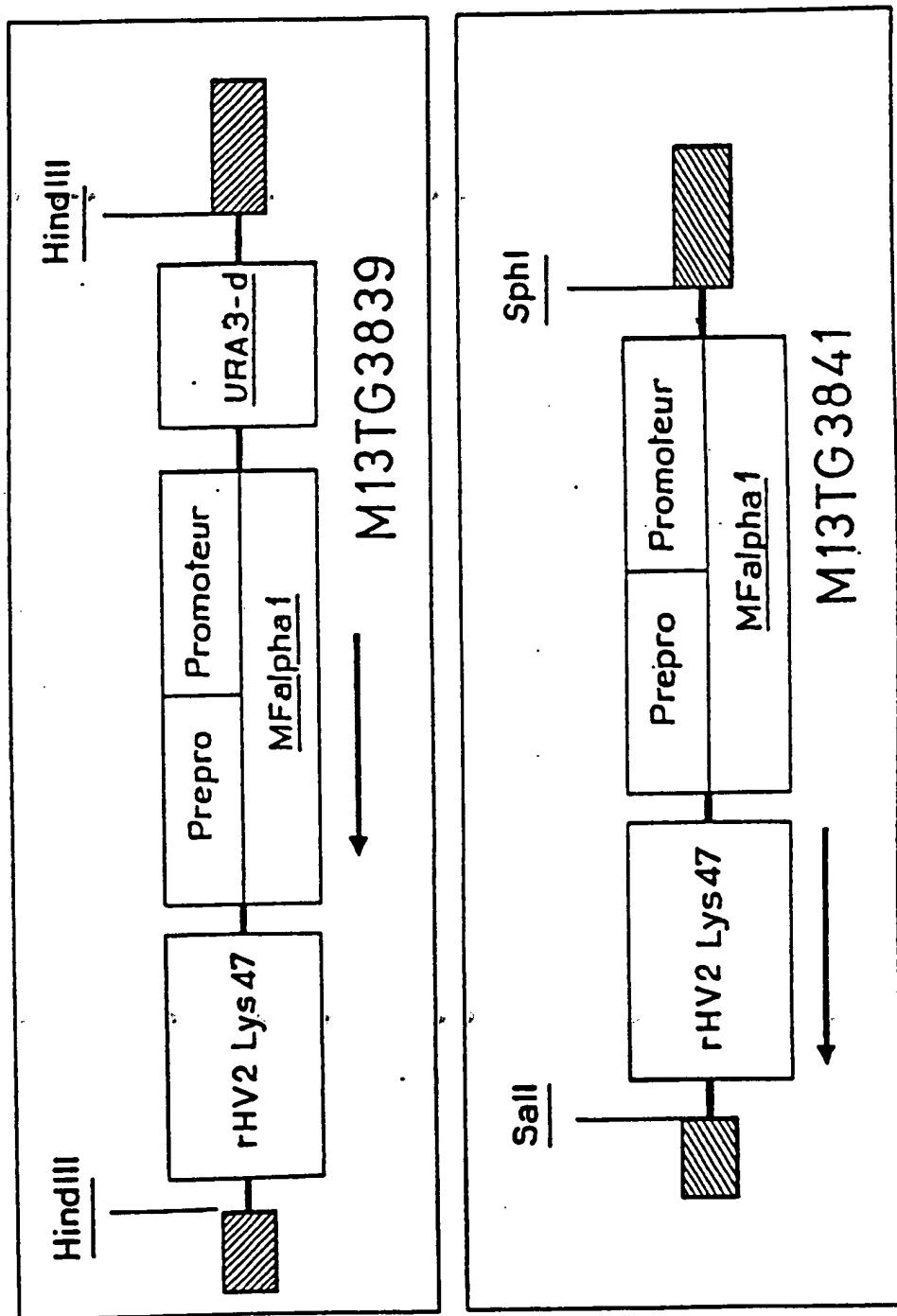
FIG. 1

FIG. 2



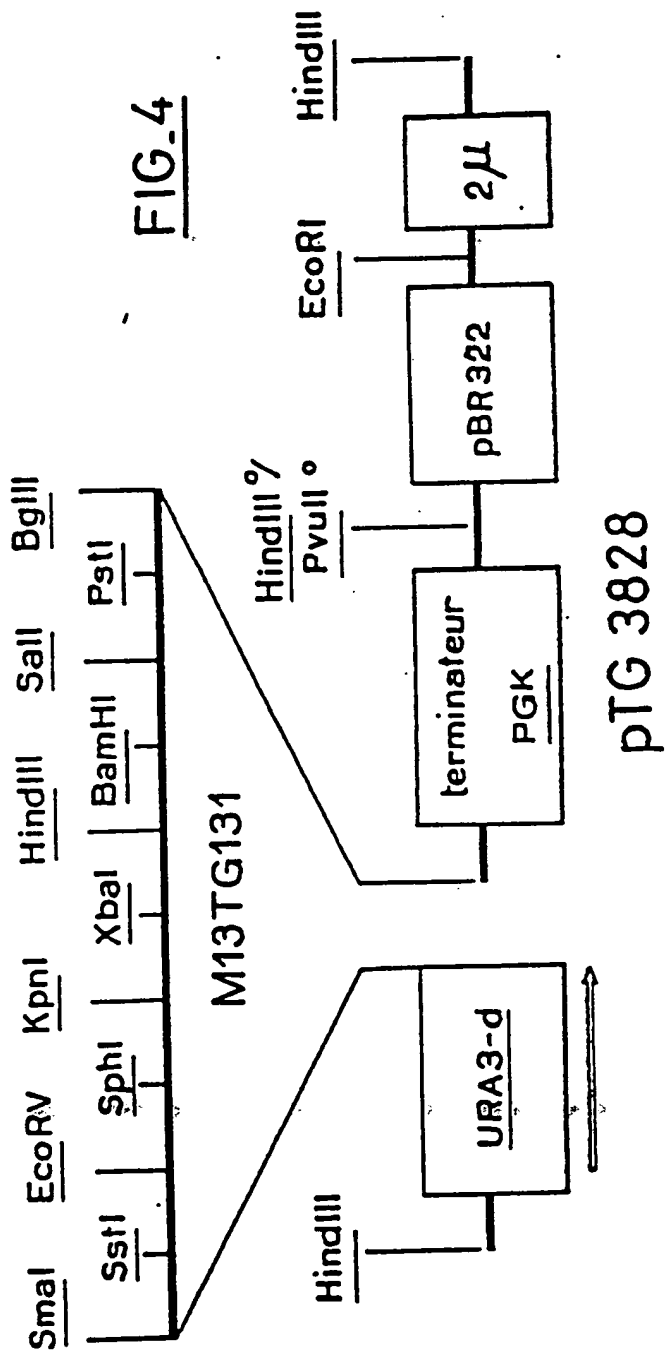
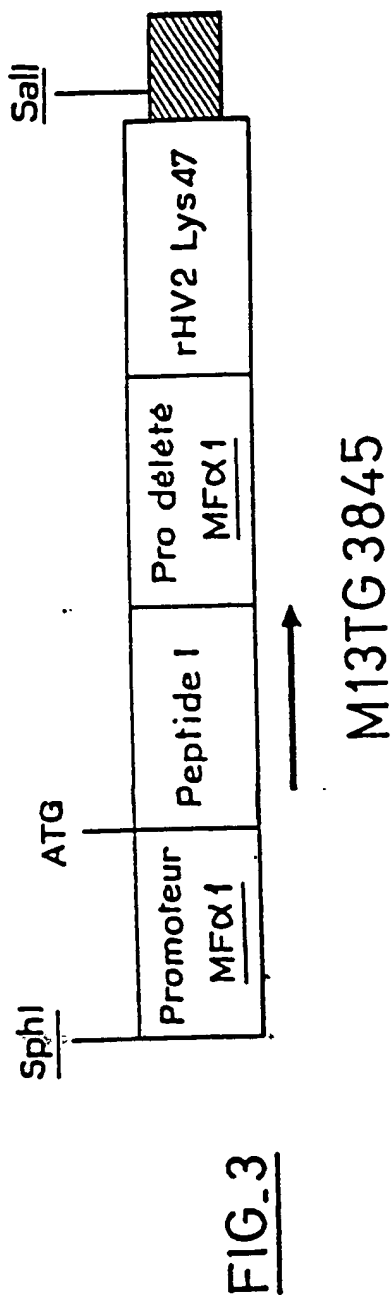
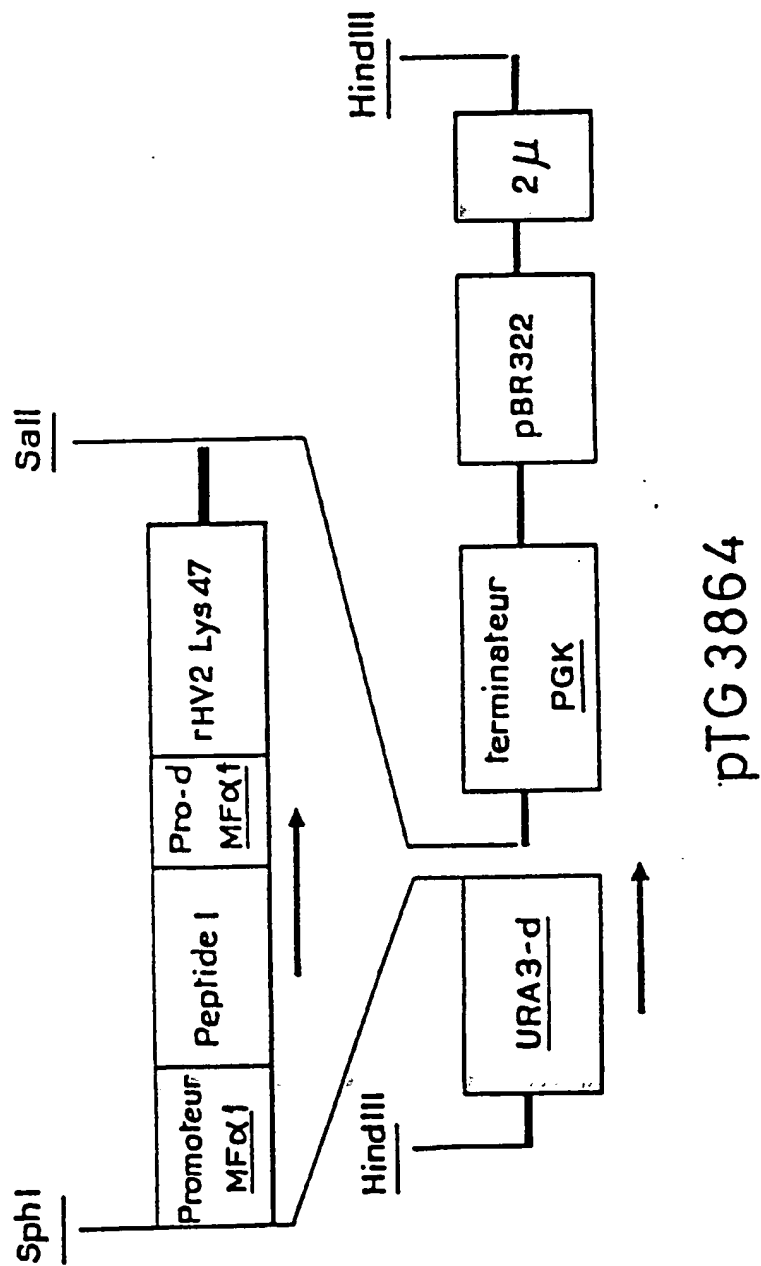


FIG. 5

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/FR 90/00306

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. <sup>5</sup> C 12 N 15/15, C 12 N 15/62, C 07 H 21/04, C 07 K 13/00, C 12 P 12/00, // A 61 K 37/02, (C 12 N 15/15, C 12 R 1:865)		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
Int.Cl. <sup>5</sup>	C 12 N 15/15, C 12 N 15/62, C 12 N 15/56, C 07 K 07/10	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT <sup>9</sup></b>		
Category <sup>10</sup>	Citation of Document, <sup>11</sup> with Indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
A	EP, A, 0252854 (TRANSGENE SA) 13 January 1988, see the whole document cited in the application ---	11-15,19-26
A	EP, A, 0273800 (TRANSGENE SA) 6 July 1988, see the whole document cited in the application ---	11-16,19-26
A	EP, A, 0158564 (TRANSGENE SA) 16 October 1985, see the whole document, in particular page 26, lines 8-19 ---	1,11,19,23-25
A	EP, A, 0225633 (CIBA-GEIGY AG) 16 June 1987, see abstract; page 17 ---	1,11-14
A	EP, A, 0200655 (TRANSGENE SA) 5 November 1986, see the whole document, in particular figure 2 cited in the application ---	1,11-14
	./-	
<p>* Special categories of cited documents: <sup>14</sup></p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
03 August 1990 (03.08.90)	25 September 1990 (25.09.90)	
International Searching Authority	Signature of Authorized Officer	
European Patent Office		

II. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	The Journal of Biological Chemistry, vol. 257, No. 24, 25 December 1982, American Society of Biological Chemists, Inc., (Baltimore, MD, US), J. Kaput et al.: "Nucleotide sequence of the yeast nuclear gene for cytochrome c peroxidase precursor", pages 15054-15058 see abstract; page 15056, figure 4	1,9
P,A	EP, A, 0349451 (CNRS) 3 January 1990 see the whole document, in particular pages 493-498 et claim 2 cited in the application	17,28,27
P,A	Journal of Bacteriology, vol. 171, No. 11, November 1989, American Society for Microbiology, (Baltimore, MD, US), F. Klebl et al.: "Molecular cloning of a cell wall exo- $\beta$ -1,3-glucanase from Saccharomyces cerevisiae", pages 6259-6264 see the whole article, in particular figure 6 cited in the application	1-8

**ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.**

FR 9000306  
SA 37090


This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 10/09/90. The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0252854	13-01-88	FR-A- 2601383	15-01-88
		FR-A- 2607516	03-06-88
		AU-A- 7536687	14-01-88
		JP-A- 1085082	30-03-89
EP-A- 0273800	06-07-88	FR-A- 2607517	03-06-88
		AU-A- 8188087	02-06-88
		JP-A- 63152987	25-06-88
		ZA-A- 8709011	27-05-88
EP-A- 0158564	16-10-85	FR-A, B 2562088	04-10-85
		FR-A, B 2569420	28-02-86
		WO-A- 8504418	10-10-85
		JP-T- 61501609	07-08-86
EP-A- 0225633	16-06-87	AU-A- 6650486	18-06-87
		JP-A- 62208296	12-09-87
EP-A- 0200655	05-11-86	FR-A, B 2593518	31-07-87
		AU-A- 5778786	18-11-86
		WO-A- 8606406	06-11-86
		FR-A- 2601383	15-01-88
		JP-T- 62502661	15-10-87
EP-A- 0349451	03-01-90	FR-A- 2633296	29-12-89
		AU-A- 3674889	04-01-90
		JP-A- 2045498	15-02-90

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale N°

PCT/FR 90/00306

<b>I. CLASSEMENT DE L'INVENTION</b> (si plusieurs symboles de classification sont applicables, les indiquer tous) *		
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB <b>CIB<sup>5</sup>:</b> C 12 N 15/15, C 12 N 15/62, C 07 H 21/04, C 07 K 13/00, C 12 P 21/00, // A 61 K 37/02, (C 12 N 15/15, C 12 R 1:865)		
<b>II. DOMAINES SUR LESQUELS LA RECHERCHE A PORTÉ</b>		
Documentation minimale consultée *		
Système de classification	Symboles de classification	
<b>CIB<sup>5</sup></b>	C 12 N 15/15, C 12 N 15/62, C 12 N 15/56, C 07 K 07/10	
Documentation consultée autre que la documentation minimale dans la mesure où de tels documents font partie des domaines sur lesquels la recherche a porté *		
<b>III. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS **</b>		
Catégorie *	Identification des documents cités, ** avec indication, si nécessaire, des passages pertinents **	N° des revendications visées **
A	EP, A, 0252854 (TRANSGENE SA) 13 janvier 1988 voir le document en entier cité dans la demande --	11-15, 19-26
A	EP, A, 0273800 (TRANSGENE SA) 6 juillet 1988 voir le document en entier cité dans la demande --	11-16, 19-26
A	EP, A, 0158564 (TRANSGENE SA) 16 octobre 1985 ... voir le document en entier, en particulier page 26, lignes 8-19 --	1, 11, 19, 23-25
A	EP, A, 0225633 (CIBA-GEIGY AG) 16 juin 1987 voir abrégé; page 17 --	1, 11-14
A	EP, A, 0200655 (TRANSGENE SA) 5 novembre 1986 ./.	1, 11-14
* Catégories spéciales de documents cités: ** « A » document définissant l'état général de la technique, non considéré comme particulièrement pertinent « E » document antérieur, mais publié à la date de dépôt international ou après cette date « L » document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée) « O » document se référant à une divulgation orale, à un usage, à une exposition ou sous autres moyens « P » document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée « T » document ultérieur publié postérieurement à la date de dépôt international ou à la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention « X » document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive « Y » document particulièrement pertinent: l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier. « & » document qui fait partie de la même famille de brevets		
<b>IV. CERTIFICATION</b>		
Date à laquelle la recherche internationale a été effectivement achevée		Date d'expédition du présent rapport de recherche internationale
3 août 1990		25.09.90
Administration chargée de la recherche internationale OFFICE EUROPEEN DES BREVETS		Signature du fonctionnaire autorisé  Nathalie Weinberg

II. DOCUMENTS CONSIDÉRÉS COMME PERTINENTS (SUITE DES RENSEIGNEMENTS INDICUÉS SUR LA DEUXIÈME FEUILLE)		
Catégorie	Identification des documents cités, avec indication, si nécessaire, des passages pertinents	N° des revendications visées
	voir le document en entier, en particulier figure 2 cité dans la demande --	
A	The Journal of Biological Chemistry, volume 257, no. 24, 25 décembre 1982, American Society of Biological Chemists, Inc., (Baltimore, MD, US), J. Kaput et al.: "Nucleotide sequence of the yeast nuclear gene for cyto- chrome c peroxidase precursor", pages 15054-15058 voir abrégé; page 15056, figure 4 --	1,9
P,A	EP, A, 0349451 (CNRS) 3 janvier 1990 voir le document en entier, en particulier pages 493-498 et revendication 2 cité dans la demande - --	17,28,27
P,A	Journal of Bacteriology, volume 171, no. 11, novembre 1989, American Society for Microbiology, (Baltimore, MD, US), F. Klebl et al.: "Molecular cloning of a cell wall exo- $\beta$ -1,3-glucanase from Saccharomyces cerevisiae", pages 6259-6264 voir l'article en entier, en particulier figure 6 cité dans la demande -----	1-8

**ANNEXE AU RAPPORT DE RECHERCHE INTERNATIONALE  
RELATIF A LA DEMANDE INTERNATIONALE NO.**

FR 9000306  
SA 37090

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche internationale visé ci-dessus.  
Lesdits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 10/09/90.  
Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets.

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
EP-A- 0252854	13-01-88	FR-A- 2601383	15-01-88
		FR-A- 2607516	03-06-88
		AU-A- 7536687	14-01-88
		JP-A- 1085082	30-03-89
EP-A- 0273800	06-07-88	FR-A- 2607517	03-06-88
		AU-A- 8188087	02-06-88
		JP-A- 63152987	25-06-88
		ZA-A- 8709011	27-05-88
EP-A- 0158564	16-10-85	FR-A, B 2562088	04-10-85
		FR-A, B 2569420	28-02-86
		WO-A- 8504418	10-10-85
		JP-T- 61501609	07-08-86
EP-A- 0225633	16-06-87	AU-A- 6650486	18-06-87
		JP-A- 62208296	12-09-87
EP-A- 0200655	05-11-86	FR-A, B 2593518	31-07-87
		AU-A- 5778786	18-11-86
		WO-A- 8606406	06-11-86
		FR-A- 2601383	15-01-88
		JP-T- 62502661	15-10-87
EP-A- 0349451	03-01-90	FR-A- 2633296	29-12-89
		AU-A- 3674889	04-01-90
		JP-A- 2045498	15-02-90

Pour tout renseignement concernant cette annexe : voir Journal Officiel de l'Office européen des brevets, No.12/82